CENTRE FOR ADVANCED INTERNET ARCHITECTURES

Real-time Internet traffic classification
Energy efficient networking

(CAIA Data Dump talk)

Dr Thuy Nguyen

Centre for Advanced Internet Architectures
Faculty of Information and Communication Technologies

About myself

1998 - 2002	Bachelor of Telecommunications Engineering, Diploma of Practice (First Class Honour), University of Technology, Sydney
2003 - 2006	PhD Student, CAIA, Swinburne University of Technology
2006 - 2008	Research Fellow, CAIA, Swinburne University of Technology
2010 - Now	Lecturer in Telecommunications Engineering Swinburne University of Technology

My Research: Past

Focused on the automation of the QoS control process: an automated, real-time IP traffic classification mechanism

Real-time Internet traffic classification

- Traditional techniques:
  - Well known, registered ports
  - Payload
- Newer approach:
  - Statistical patterns in externally observable attributes of the traffic
  - Application of Machine Learning techniques
- Problem:
  - Relied on bi-directional, full-flow statistics
  - Assumed explicit direction implied by the first packet captured, or a known client-server relationship
  - Some used the first few packets of a flow

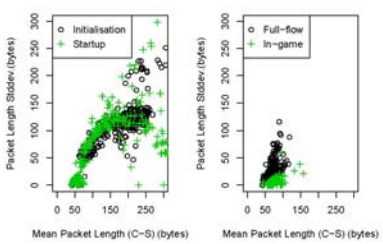
Real-time Internet traffic classification

- Reality:
  - Real-time classification, well before a flow has finished
  - Classifiers may miss an arbitrary number of packets from the start of a flow, and
  - Be unsure of the direction in which the flow started.

- My proposal: classification using a sliding window of the most recent N packets of a flow

Real-time Internet traffic classification

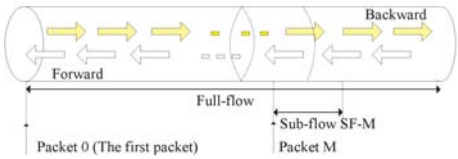
- Challenges:
  - a target application's short-term traffic statistics vary within the lifetime of a single flow
  - most Internet applications exhibit asymmetric traffic characteristics in the client-to-server and server-to-client directions



Packet Length from Client to Server **Packet Length from Server to Client**

Real-time Internet traffic classification

- Solution:
  - training a ML classifier on a set of multiple short “sub-flows” - a collection of N consecutive packets extracted from full-flow samples
  - sub-flows are picked from regions of the application’s flow with noticeably different statistical characteristics
  - a novel use of unsupervised ML algorithms for the automated selection of appropriate sub-flows
  - synthesising a complementary version of every sub-flow in the reverse direction



Full-flow vs. Sub-flow

Real-time Internet traffic classification

- Results:
  - Naive Bayes and C4.5 Decision Tree ML algorithms, identification of interactive online game – Wolfenstein Enemy Territory (ET) and VoIP traffic.
  - Sliding window of 25 packets:
    - ET: 98.9% - 99.3% Recall and 87% - 97% Precision
    - VoIP: 95.7% - 100% Recall and 95.4% - 99.2% Precision
  - Classification performance is maintained:
    - when the classification is initiated at an arbitrary point within a flow, and
    - independent of the direction of the first packet captured.
- Utilised in:
  - ANGEL - Automated Network Games Enhancement Layer part of the Smart Networks project of the Smart Internet Technology CRC
    http://caia.swin.edu.au/sitcrc/angel/
  - DIstributed Firewall and Flow-shaper Using Statistical Evidence (DIFFUSE) project
    http://caia.swin.edu.au/urp/diffuse

My Research: Present

Focused on Green ICT: energy efficiency in content distribution, data centres and cloud computing.

Energy efficiency for data centres & cloud computing

- Energy consumption is a major concern:
  - Power is the second-highest operating cost in 70% of all data centres
  - Data centres are responsible for 10s millions of metric tons of CO2 emissions annually, more than 5% of the total global emissions.
- A possible way to save energy:
  - Run a server at (or near) full utilisation
  - Turn off idle ones
- Server work load:
  - An idle server may still use up to 60% of its peak power
  - Rarely completely idle or fully utilised: mostly between 10 – 50% of maximum utilization.

Energy efficiency for data centres & cloud computing

- Virtualisation:
  - Consolidation of hardware and workload
  - Better utilisation of hardware → potentially savings on energy
- Challenges:
  - Energy efficiency in the consolidation/migration process
    - E.g. minimising the total migration time, so that the original node could be turned off asap.
  - Energy efficiency in communication protocols
    - E.g. migration between different sub-nets, or over wide area networks

Energy efficiency for content distribution

- Peer to peer (P2P)
- Content Distribution Network (CDN)
- ISP assisted P2P and CDN for more efficient content distribution?